JOURNAL OF VIROLOGY, May 2006, p. 5050–5058 Vol. 80, No. 10
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.10.5050–5058.2006

X-Ray Crystallographic Structure of the Norwalk Virus Protease at 1.5-Å Resolution

Corinne E. Zeitler,[1] Mary K. Estes,[1] and B. V. Venkataram Prasad[2]*

Department of Molecular Virology and Microbiology[1] and Verna and Marrs McLean Department of Biochemistry and Molecular Biology,[2] Baylor College of Medicine, Houston, Texas 77030

Received 18 December 2005/Accepted 24 February 2006

Norwalk virus (NV), a member of the *Caliciviridae* family, is the major cause of acute, epidemic, viral gastroenteritis. The NV genome is a positive sense, single-stranded RNA that encodes three open reading frames (ORFs). The first ORF produces a polyprotein that is processed by the viral cysteine protease into six nonstructural proteins. We have determined the structure of the NV protease to 1.5 and 2.2 Å from crystals grown in the absence or presence, respectively, of the protease inhibitor AEBSF [4-(2-aminoethyl)-benzene-sulfonyl fluoride]. The protease, which crystallizes as a stable dimer, exhibits a two-domain structure similar to those of other viral cysteine proteases with a catalytic triad composed of His 30, Glu 54, and Cys 139. The native structure of the protease reveals strong hydrogen bond interactions between His 30 and Glu 54, in the favorable *syn* configuration, indicating a role of Glu 54 during proteolysis. Mutation of this residue to Ala abolished the protease activity, in a fluorogenic peptide substrate assay, further substantiating the role of Glu 54 during proteolysis. These observations contrast with the suggestion, from a previous study of another norovirus protease, that this residue may not have a prominent role in proteolysis (K. Nakamura, Y. Someya, T. Kumasaka, G. Ueno, M. Yamamoto, T. Sato, N. Takeda, T. Miyamura, and N. Tanaka, J. Virol. 79:13685–13693, 2005). In the structure from crystals grown in the presence of AEBSF, Glu 54 undergoes a conformational change leading to disruption of the hydrogen bond interactions with His 30. Since AEBSF was not apparent in the electron density map, it is possible that these conformational changes are due to subtle changes in pH caused by its addition during crystallization.

Norwalk virus is a member of the *Norovirus* genus of the viral family *Caliciviridae*. Noroviruses are the major causative agents of nonbacterial, acute gastroenteritis in humans. Currently, understanding the replication of noroviruses is hindered due in part to the lack of a cell culture system or an animal model. The Norwalk virus genome is a 7.7-kb positive sense, single-stranded RNA that encodes three open reading frames (ORFs). ORFs 2 and 3 encode the viral capsid protein, VP1 (19), and the minor structural protein, VP2 (17), respectively. A 200-kDa polyprotein is encoded by ORF 1, and this polyprotein is cleaved by the viral protease into six nonstructural proteins necessary for replication (5, 9). The six nonstructural proteins include an RNA-dependent RNA polymerase (30); the viral protease (38); the genome-linked viral protein (VPg), which may act as an mRNA cap (13); p22, which shares amino acid motifs with the picornavirus 3A protein (9); p41, which has NTPase activity and a helicase domain (35); and p48, which interferes with cellular trafficking (16). These proteins were identified originally through sequence motif similarity to proteins from the *Picornaviridae* family, which also encode a large polyprotein that is cleaved into the nonstructural proteins necessary for replication by the viral 3C and 2A proteases.

Similar to the case in picornaviruses, the Norwalk virus protease (NV^{PRO}) is necessary to cleave its viral polyprotein into the six nonstructural proteins (9, 18, 39, 40) required for viral maturation and replication. The X-ray structures of three related picornaviral 3C proteases show a chymotrypsin-like (42) fold with a cysteine as the active-site nucleophile (1, 37). These viral cysteine proteases, such as the well-characterized picornaviral 3C and 2A proteases, contain a catalytic triad that is composed of the active-site cysteine, a histidine, and either a glutamate or aspartate. Therefore, the 3C and 3C-like proteases are considered to be a distinct group of serine-like (chymotrypsin-like) proteases and are different from cysteine (papain-like) proteases (3, 4, 31). The viral cysteine proteases use a proteolytic mechanism similar to that of the serine proteases with a cysteine sulfhydryl group acting as the nucleophile instead of the serine hydroxyl group. In both cases, the formation of the tetrahedral transition state intermediate is facilitated by hydrogen bonds between the oxyanion and two main chain amine groups. Typically, the glutamate or aspartate residue is used to orient the imidazole ring of histidine in order to neutralize the charge that develops during this tetrahedral transition state formation.

The crystal structures of poliovirus (PoV) (27), human rhinovirus (HRV) (23, 24), hepatitis A virus (HAV) (43), and foot-and-mouth disease virus (7) 3C proteases have revealed a similar catalytic triad in the active site with a cysteine as the nucleophile, histidine as the general base, and glutamate or aspartate as an anion that correctly positions the imidazole ring of the active site histidine. Site-directed mutagenesis was used to determine that the catalytic triad of the 3C-like protease from rabbit hemorrhagic disease virus, a member of the *Caliciviridae* (10, 41), is composed of a cysteine, a histidine, and an aspartate. Previously, it has been shown through mutational analysis and polyprotein cleavage experiments that Cys 139

* Corresponding author. Mailing address: Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Alkek Building N410, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Phone: (713) 798-5686. Fax: (713) 798-1625. E-mail: vprasad@bcm.tmc.edu.

and His 30 are members of the Norwalk virus protease catalytic triad (18, 39). However, the identity of the acid component of the catalytic triad has remained controversial. Multiple sequence alignments between noroviruses and picornaviruses were used in an attempt to determine the active-site anion (39). Based on this analysis, two possible anions were identified: Glu 54, which corresponds to the active site aspartate of the rabbit hemorrhagic disease virus protease, or Asp 67, which corresponds to active-site glutamate of PoV and HRV 3C proteases (39). Mutational studies of Norwalk virus by Hardy et al. suggest that Glu 54 is required for proteolysis of the polyprotein (18). However, based on the mutational analysis by Someya et al. and recent crystallographic analysis at 2.8-Å resolution, Nakamura et al. have suggested that a catalytic dyad, and not a triad, is responsible for catalysis in the Chiba virus (a related norovirus) protease (29, 39), similar to the mechanism of the coronavirus main protease (2), another viral cysteine protease.

To investigate the catalytic mechanism of norovirus proteases, we determined the structure of the NV^{PRO} to 1.5-Å resolution. Based on our structural and biochemical analysis, we report here that Glu 54 is, in fact, the anion of the catalytic triad since it acts to position His 30 in the catalytic site through strong hydrogen bonding interactions and is required for proteolysis of a fluorogenic peptide substrate. Furthermore, we have determined a 2.2-Å structure of the NV^{PRO} from the crystals grown in the presence of one of its known inhibitors, AEBSF [4-(2-aminoethyl)-benzenesulfonyl fluoride], in order to investigate the mechanism of inhibition (9). In this structure, we have found a conformational change in the residues of the catalytic triad, which results in the loss of the hydrogen bond interactions between the imidazole ring of the active-site histidine and the active-site glutamate.

MATERIALS AND METHODS

Protease expression and purification. Full-length NV^{PRO} was cloned into the expression vector pET41 Ek/LIC with an additional thrombin cleavage site and expressed in *Escherichia coli* BL21(DE3) in minimal media in the presence of seleno-methionine (Se-Met). The glutathione *S*-transferase (GST)-tagged Se-Met NV^{PRO} was purified over a glutathione-Sepharose column and the GST tag was removed with thrombin. Size exclusion chromatography was performed to separate the GST tag from NV^{PRO}. The protease inhibitor phenylmethylsulfonyl fluoride (PMSF) was present throughout purification. Se-Met NV^{PRO} was concentrated and stored in 25 mM Tris-HCl (pH 8.0)–100 mM NaCl–5 mM dithiothreitol, crystallized, and used for multiple anomalous dispersion (MAD) experiments.

Another full-length NV^{PRO} clone was constructed with an amino-terminal His_6 tag in the expression vector pET46 Ek/LIC and expressed in BL-21(DE3) *E. coli*. This NV^{PRO} was purified over a nickel affinity column and dialyzed overnight into 50 mM NaH_2PO_4 (pH 4.7)–100 mM NaCl–10 mM β-mercaptoethanol.

Site-directed mutagenesis. Glu 54 was mutated to Ala (E54A) by using the Stratagene QuikChange mutagenesis kit, according the manufacturer's protocol, using the primer 5′-GCAATCCACCAAGCAGGTGCGTTCACACAATTCAG GTTC-3′ and its reverse complement. The resulting clones were completely sequenced to confirm the mutation and to verify that no additional mutations were present. Mass spectroscopic analysis was performed to confirm the mutation of Glu 54 to Ala in the purified NV^{PRO}.

Activity assay for purified native NV^{PRO} and E54A mutant NV^{PRO}. For activity assays, purified His-tagged NV^{PRO} was dialyzed into 50 mM NaH_2PO_4 (pH 8)–50 mM NaCl–1 mM DTT. The protease was mixed with a fluorogenic peptide substrate (Edans-EPDFHLQGPEDLAK-Dabcyl), corresponding to the cleavage site between p48 and p41 in the polyprotein, which leads to an increase in fluorescence when the peptide is cleaved. The assay was monitored by using a fluorescent plate reader at excitation/emission wavelengths of 360 and 460 nm, and there was no increase in fluorescence intensity in the absence of NV^{PRO} (Fig. 1B). In addition, the E54A mutant NV^{PRO} was purified as was the wild-type His-tagged NV^{PRO} and was assayed under the same conditions with the same fluorogenic peptide substrate. The assays for the wild-type and mutant NV^{PRO} were performed a total of eight times.

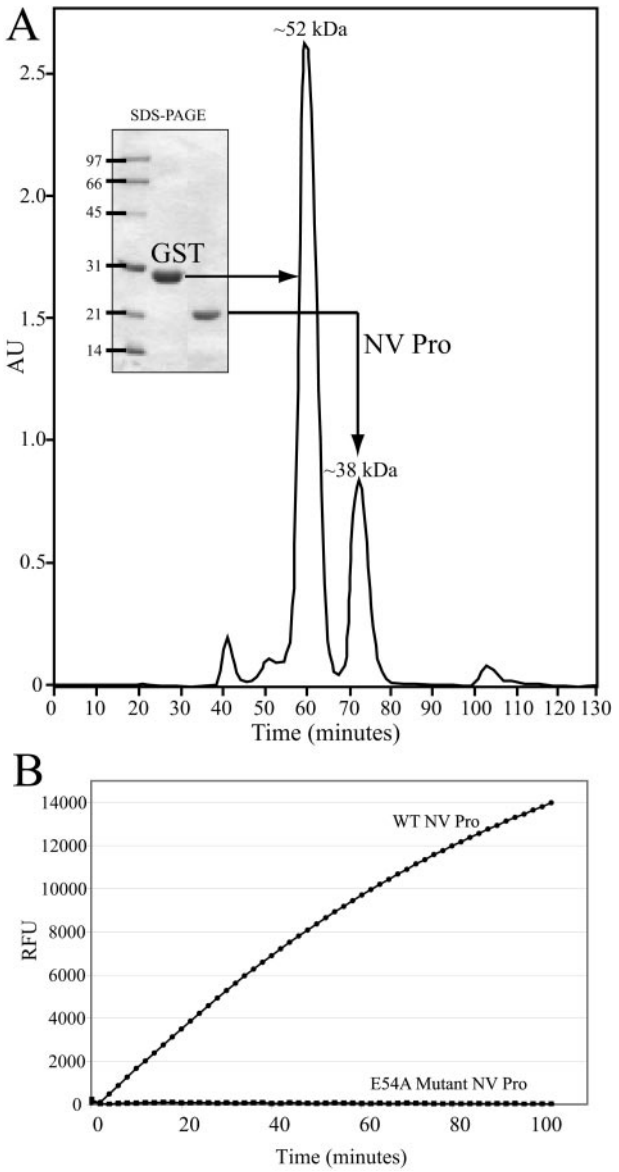

FIG. 1. Purification and activity of NV^{PRO}. (A) GST-NV^{PRO} was purified by size exclusion chromatography after thrombin cleavage, and the protease was observed to form a dimer. The size exclusion profile of the His-tagged NV^{PRO} appears identical with the same peak at ~38 kDa. The inset is a Coomassie blue-stained sodium dodecyl sulfate-polyacrylamide gel electrophoresis of molecular weight markers the indicated fractions from the peaks. (B) Activity of His-tagged NV^{PRO} after purification as shown by mixing the protease with the fluorogenic peptide substrate Edans-EPDFHLQGPEDLAK-Dabcyl, where fluorescence intensity increases with cleavage of the peptide between QG. The E54A mutant was assayed under the same conditions and is unable to cleave the fluorogenic peptide substrate. Shown on the *y*-axis are relative fluorescence units (RFU), measured at excitation/emission wavelengths of 360/460 nm. The experiments were repeated a total of eight times, and the error bars cannot be seen on the graph due to the scale on the *y* axis and the low standard deviation between each trial.

TABLE 1. Data collection and refinement statistics

Parameter[a]	Results for crystal		
	Se-Met	Native	AEBSF
Data collection statistics			
Wavelength (Å)	0.97857 (peak) 0.97948 (inflection) 0.97964 (remote)	0.9794	0.9794
Resolution limit (Å)	30.0–3.15 (3.26–3.15)	31.5–1.5 (1.55–1.50)	39.44–2.2 (2.28–2.20)
Total reflections	298,028	507,583	152,566
Unique reflections	20,054	36,055	11,542
Completeness	99.9 (100.0)	99.1 (97.9)	99.6 (98.8)
Redundancy	14.3 (15.1)	14.08 (13.83)	13.22 (13.01)
R_{merge}*	0.093 (0.494)	0.063 (0.457)	0.085 (0.326)
$I/\sigma(I)$	7.5 (2.1)	16.7 (4.1)	16.8 (5.9)
Cell parameters			
Space group	$P6_1$	$P6_522$	$P6_522$
a and *b* (Å)	130.462	121.833	120.490
c (Å)	119.889	51.369	51.249
α and β (°)	90.0	90.0	90.0
γ (°)	120.0	120.0	120.0
Refinement statistics			
Resolution range (Å)	30–3.1	30–1.50	30–2.2
R_{work} (%)	27.7	18.5	21.1
R_{free} (%)	33.1	21.1	24.6
RMS deviation from ideal geometry			
Bonds (Å)	0.014	0.008	0.006
Angles (°)	1.607	1.207	1.253
Avg B factor (Å^2)	78.208	17.310	24.958
Ramachandran plot (%)			
Favored	84.1	92.6	90.4
Allowed	14.5	7.4	9.6
Generously allowed	0.7	0	0
Disallowed	0.7	0	0

[a] *$R_{merge} = (\Sigma_h\Sigma_j|I_j(h) - <I_h>|)/\Sigma_h\Sigma_j I_j(h)$, where $I_j(h)$ is the *j*th measurement of the intensity of reflection *h* and $<I(h)$ is the mean intensity of symmetry-equivalent reflections. R_{free} is the *R* factor for a subset (10%) of the reflections that was selected prior to refinement calculations and was not included in the refinement. Ramachandran plots were generated by the program PROCHECK (22). Numbers in parentheses indicate the highest resolution shell.

Crystallization. Se-Met NV^{PRO} and native NV^{PRO} were crystallized either by the hanging drop vapor diffusion or sitting drop vapor diffusion methods at room temperature. The best Se-Met NV^{PRO} (25 mM Tris-HCl [pH 8.0], 100 mM NaCl, 5 mM DTT) crystals were obtained with sitting drops in 1.4 M sodium potassium phosphate with 0.1 M lithium chloride as an additive. The cryoprotectant was the same as the crystallization condition with 25% polyethylene glycol (PEG). The best native His-tagged NV^{PRO} crystals were obtained by the sitting drop method with a well solution of 0.1 M lithium sulfate–0.1 M bis-Tris (pH 5.5)–20% PEG 3350–0.1 M EDTA. The native His-tagged NV^{PRO} (50 mM NaH_2PO_4 [pH 4.7], 100 mM NaCl, 10 mM β-mercaptoethanol) was also cocrystallized in the presence of the inhibitor AEBSF. AEBSF-NV^{PRO} crystals were obtained by hanging-drop vapor diffusion in 0.2 M magnesium chloride–0.1 M bis-Tris (pH 5.5)–30% PEG 3350 with 1 μl of 1 M AEBSF, prepared in 50 mM NaH_2PO_4–50 mM NaCl added to the drops for a final concentration of 0.1 M. All crystal trays were set up with a protease concentration of 2 mg/ml.

Data collection. All data were collected at Argonne National Labs' Advanced Photon Source, beamline SBC-19ID. MAD diffraction data to 3.1 Å was collected for the Se-Met NV^{PRO} crystals. The wavelengths used for MAD were as follows: 0.97948 Å for peak, 0.97964 Å for inflection, and 0.97964 Å for remote data. The data integration and scaling were performed with HKL2000 (32). Diffraction data for His-tagged NV^{PRO} crystals were obtained to 1.5 Å (native) and to 2.2 Å (grown in the presence of AEBSF) using a wavelength of 0.9794 Å. The data integration and scaling were performed with the program D*Trek (36).

The Se-Met NV^{PRO} crystallized in a space group of $P6_1$, whereas the His-tagged NV^{PRO} crystallized in the space group $P6_522$. The data collection statistics are summarized in Table 1.

Structure determination, refinement, and model building. The positions of the Se atoms were identified using *SnB* (26). Refinement of the Se atom positions and electron density map calculations were performed with SHARP (11). Residues were positioned manually using the program O (21) based on two F_o-F_c and F_o-F_c maps, followed by iterative cycles of model building with O (21) and refinement with CNS (12).

The 1.5- and 2.2-Å structures were determined by molecular replacement with the Se-Met NV^{PRO} structure using Phaser (25). The atomic models were obtained using ARP/wARP (33), followed by iterative cycles of refinement with Refmac5 (28) (1.5-Å data) or CNS (2.2-Å data) (12). Waters were allocated to peaks >3σ in the F_o-F_c difference map using COOT (15). During the course of the refinement and after the final refinement, the stereochemistry of both the structures was checked by using PROCHECK (22). Final refinement statistics are shown in Table 1. Structural figures were rendered with PyMOL (14). Buried surface areas for the dimer complexes were calculated with CNS (12).

Protein structure accession numbers. The coordinates and associated structure factors have been deposited into the Protein Data Bank (PDB code 2FYQ and 2FYR for the 1.5-Å structure and the 2.2-Å structure, respectively).

RESULTS AND DISCUSSION

NV^{PRO} is a dimer in solution. In these studies, the NV^{PRO} was purified by expressing two different constructs. In the initial stages, based on the previous successful expression of the functionally active NV protease (9), a GST construct was used. The native and Se-Met protein, using this construct, yielded crystals that only diffracted to 3.1 Å. The data from the Se-Met NV^{PRO} crystals were adequate to obtain an initial model of the

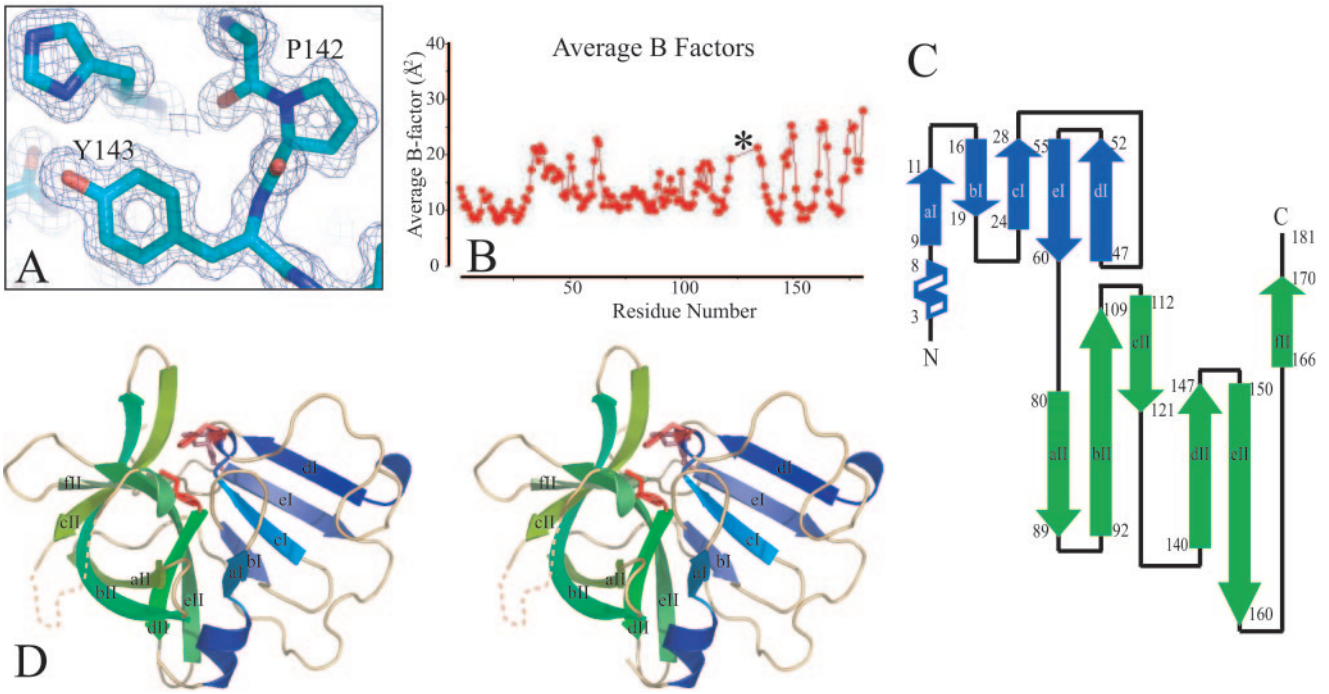

FIG. 2. Structure of the Norwalk virus protease at 1.5-Å resolution. (A) Representative sample of the electron density map with modeled amino acids. (B) Average B factors for each residue. The disordered loop corresponding to amino acids 122 to 133 is labeled with an asterisk. The average B-factor for the entire protease monomer is 17.3 Å^2. B-factors for the catalytic residues are also low, at 12.3 Å^2 (H30), 12.49 Å^2 (E54), and 11.52 Å^2 (C139). (C) Secondary structure topology diagram. Amino acid numbers for each of the strands are shown, and the strands are labeled for each domain that corresponds to the labels in panel D. (D) The Norwalk virus protease consists of two domains separated by a cleft that contains the active-site catalytic triad. The triad (shown in red) consists of His 30, Glu 54, and Cys 139. Domain I (blue) has a short amino-terminal α-helix, followed by a twisted β-sheet and domain II (green) is a six-stranded β-barrel.

NV^{PRO}. Subsequently, a His-tagged construct was used that yielded crystals that diffracted to significantly higher resolution, and these data were phased by using the low-resolution model. During the purification of the GST construct of NV^{PRO}, after thrombin cleavage, by size exclusion chromatography, it was noted that the peak representing the cleaved NV^{PRO} corresponded to twice the molecular weight of the protein, indicating thereby that recombinant NV^{PRO} exists as dimers in solution (Fig. 1A). A similar profile was observed for the His-tagged construct, indicating that NV^{PRO} is a dimer in solution irrespective of expression and purification protocols. The activity of the purified His-tagged NV^{PRO} was ascertained by using a fluorogenic peptide substrate assay (Fig. 1B). This type of assay uses fluorescent resonance energy transfer to detect cleavage of the peptide, where there is a fluorescent donor and a nonfluorescent acceptor on the peptide. The fluorescence of the peptide remains quenched until the peptide is cleaved and the fluorescent donor is no longer in contact with the acceptor. The increase in fluorescence was measured in relative fluorescence units to determine the activity of NV^{PRO}.

Overall structure of the Norwalk virus protease. The quality of the electron density map of the native NV^{PRO} at 1.5 Å is excellent, with an overall B-factor of 17.3 Å^2, and allowed us to model the positions of the main chain, side chain atoms, and several solvent molecules unambiguously (Fig. 2A and B). The NV^{PRO} adopts a serine protease-like fold that consists of two β-barrel domains separated by a cleft within which lie the active-site catalytic residues (Fig. 2C and D). NV^{PRO} domain I begins with a two turn α-helix (amino acids [aa] 3 to 8), followed by five β-strands (aI to dI) connected by large loops that create a twisted antiparallel β-sheet. Domain I is connected to domain II via a large loop (aa 61 to 79). Domain II is defined by a six-stranded (aII to fII) antiparallel β-barrel consisting of residues 80 to 170. The strands in domain II are connected by smaller loops, where loop 123-133 is disordered and not visible in the electron density map. The active site is located at the opening of the cleft between domains I and II composed of residues His 30, Glu 54, and Cys 139. His 30 and Glu 54 are located on two separate loops that connect strands cI to dI and dI to eI, respectively. Cys 139 is part of the large loop that connects strands cII to dII in domain II. The conformation of the loops and β-strands in domains I and II are likely to play a role in positioning the active-site residues for proteolysis. Although the NV^{PRO} structure generally resembles other viral cysteine proteases, such as poliovirus (27), HAV (6), HRV (24), and foot-and-mouth disease virus (7), where both domains are β-barrels, one distinct difference is that NV^{PRO} has only one β-barrel domain while the other is a twisted β-sheet.

Protease dimers. Based on the size exclusion chromatography results, NV^{PRO} forms dimers in solution (Fig. 1A). Is it a dimer in the crystal? The His-tagged NV^{PRO} crystallized in the $P6_522$ space group with only one protease molecule per asymmetric unit. However, when crystallographic symmetry operations were applied, a distinct tail-to-tail dimer, formed using a crystallographic twofold symmetry, was repeated throughout crystal (Fig. 3A). The total buried surface area for this dimer is

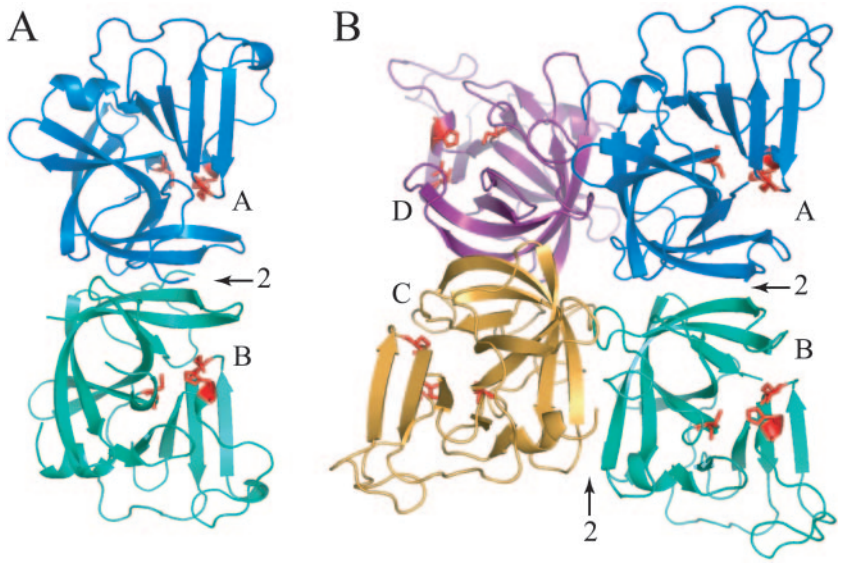

FIG. 3. NV^{PRO} is a dimer in solution, as well as in crystals. The NV^{PRO} is observed as a dimer during size exclusion chromatography, as well as in the crystal. (A) The $P6_522$ asymmetric unit contains one protease molecule. When symmetry operations are applied, the dimer shown is repeated throughout the crystal. The crystallographic twofold axis is shown with an arrow. (B) There are four molecules in the $P6_1$ asymmetric unit of the Se-Met NV^{PRO} crystals. The noncrystallographic twofold axes are shown with arrows. The AB and CD dimers correspond to the dimer seen in the high-resolution structure. In both figures, the catalytic residues are shown in red.

~2,380 Å^2, a value that is consistent with other homodimers (20). In the Se-Met NV^{PRO} crystal, which has four molecules in the asymmetric unit, two dimers, related by a local twofold axis, juxtaposed with each other can be identified with a very similar subunit orientation and interface as the dimer in the His-tagged crystals (Fig. 3B). A similar organization of the four molecules in the asymmetric unit is seen in the recently determined structure of Chiba virus protease (29). However, the buried surface areas for these dimers (~1,400 Å^2) are smaller because one of the polypeptide segments in the interface is disordered. Although the RMS deviation (for the matching C_α atoms) between the monomers in the two different crystals forms is ~0.7 Å, the RMS deviation between the dimers is ~2.8 Å, indicating that the orientation between the subunits of the dimers in the Se-Met NV^{PRO} crystal is slightly altered, perhaps influenced by the crystallization conditions and crystal packing forces.

The β-barrel of each NV^{PRO} monomer interacts along the crystallographic twofold axis to form the dimer, which is stabilized by both van der Waals and hydrogen bond interactions involving four regions on each monomer: Glu 79 that is directly before the β-strand aII; Ser 106 and Arg 108 on β-strand bII; Ser 163 and Thr 166 on β-strand fII; and Glu 177 and Glu 181, which are part of the C-terminal loop that is disordered in the Se-Met NV^{PRO} structure. These residues are conserved among genogroup I noroviruses and have conservative substitutions in the other norovirus genogroups, suggesting that formation of dimers may be a common feature of norovirus proteases. Viral cysteine proteases existing as dimers have been reported for the HAV 3C protease, which has a higher affinity for RNA when dimerized (34), and the SARS coronavirus main protease, which has two domains similar to other viral cysteine proteases and a third domain that interacts to form dimers (2). Further mutational and functional studies are required to establish the role of dimerization of NV^{PRO} in the replication of Norwalk virus.

Catalytic triad. Based on the structural comparison of our NV^{PRO} structure with other serine proteases and viral cysteine proteases, along with previously published studies identifying the catalytic residues of NV^{PRO} as His 30, Cys 139, and possibly Glu 54 (18, 39), the residues important for enzymatic activity could be located. Serine and viral cysteine proteases use a triad of residues to cleave the target peptide bond. Common serine proteases have Ser, His, and Asp residues in their catalytic triad, whereas in viral cysteine proteases have replaced Ser with a Cys residue, His, and either Glu or Asp at the third position. In the NV^{PRO} structure, these residues correspond to Cys 139, His 30, and Glu 54, which function as a nucleophile, general base, and anion, respectively.

There has been speculation about the requirements for an anion in the NV^{PRO} and which residue might act as the anion, if any. Based on multiple sequence alignments, Someya et al. initially suggested Glu 54 or Asp 67 as candidate anions for noroviruses; however, their mutational studies, using a fusion protein as a substrate, did not show the requirement of either (37). In contrast, mutational analysis of NV^{PRO} by Hardy et al. suggested that Glu 54 is indeed the anion and may be required for proteolysis (18). The 1.5-Å NV^{PRO} structure, shown here, clearly demonstrates that Glu 54 is used to orient the imidazole ring of His 30 for proteolysis. Glu 54 is hydrogen bonded to the imidazole ring of His 30 (Oε1 to Nδ1, 2.7 Å), to the backbone nitrogen of His 30 (Oε2 to N, 2.9 Å), and to a water molecule (Oε2, 2.7 Å) (Fig. 4A and Fig. 5, red). The network of strong hydrogen bond interactions involving the anion, the general base, and the water molecule is very similar to that seen in serine proteases in which the role of the anion is definitively established. However, the water molecule in the NV^{PRO} structure replaces the hydroxyl group of a serine residue used in

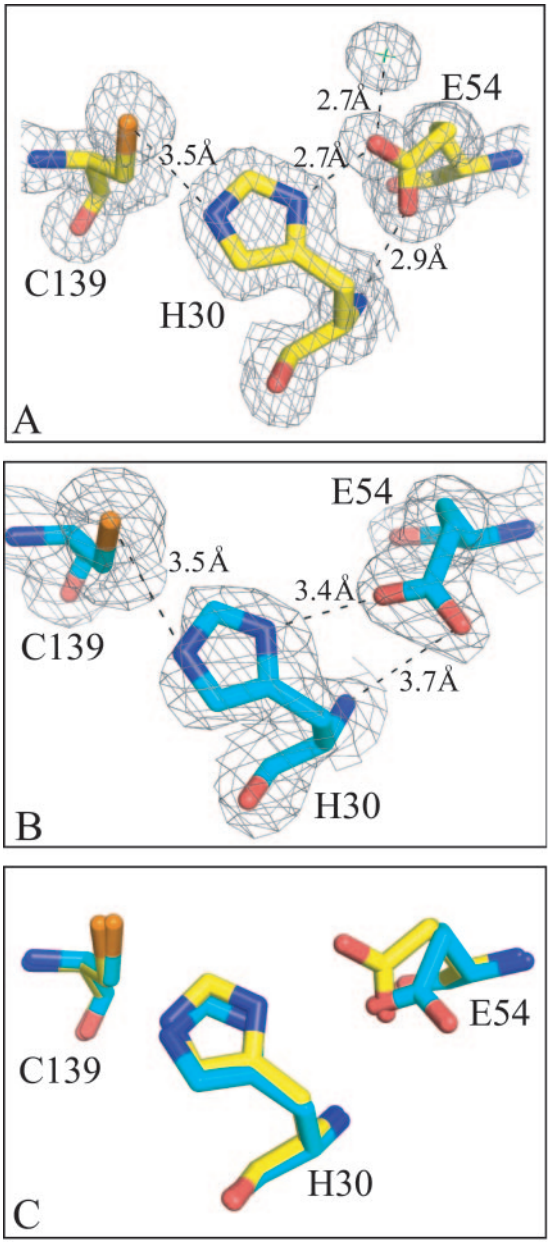

FIG. 4. Catalytic residues undergo conformational change in presence of AEBSF. (A) Catalytic residues of the native NV^{PRO} at 1.5-Å resolution inside of the 2Fo-Fc map. (B) Catalytic residues at 2.2-Å resolution inside of the 2Fo-Fc map, from crystals grown in the presence of AEBSF. (C) Comparison of native triad and the triad from the crystals grown in the presence of AEBSF. When NV^{PRO} is grown in the presence of AEBSF, the stabilizing Glu-His hydrogen bonds are broken. In all of the figures, the 1.5-Å structure is represented in yellow, the 2.2-Å structure is cyan, the oxygen atoms are red, and the nitrogen atoms are blue and sulfur is orange. Hydrogen bonds are represented by dashed lines and distances are marked in angstroms.

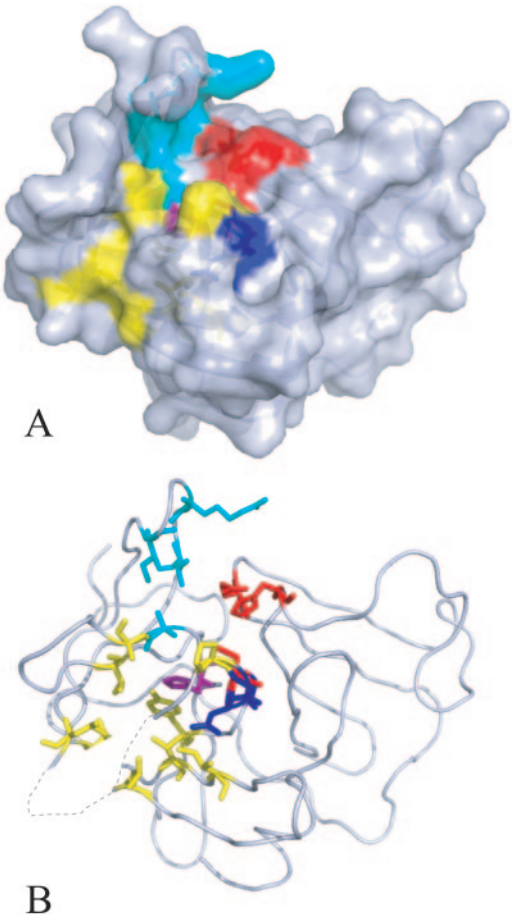

FIG. 5. Substrate binding sites on the surface of NV^{PRO}. (A) The 1.5-Å NV^{PRO} structure is shown as a ribbon representation, with the substrate binding residues shown as sticks and color coded as described below. (B) The surface representation of the 1.5-Å structure of the NV^{PRO}. In both figures, the S1 specificity pocket is shown in yellow with His 157 highlighted in magenta. The S2 pocket is colored cyan, and the oxyanion hole is dark blue. The catalytic residues are shown in red. The amino terminus and carboxyl terminus are labeled with an N and a C, respectively.

serine proteases. Such an interaction between Glu 54 and His 30, referred to as a *syn* configuration, is more favorable and is associated with enhanced catalytic efficiency (7). In addition, mutation of Glu 54 to Ala results in a loss of proteolytic activity as assayed using a fluorogenic peptide substrate (Fig. 1B), which highlights the requirement of Glu 54 during proteolysis.

The contrasting results between the NV and Chiba virus proteases, using a triad versus a dyad, respectively, is surprising considering that both of these viruses are from the same genogroup, with 91% aa identity in the protease sequence, and that Glu 54 is conserved throughout. It should be emphasized here that a similar dilemma occurred with the HAV protease, in which a dyad instead of triad was suggested for its activity in early structural studies. However, recent high resolution crystallographic studies have shown a triad with aspartate in the *syn* configuration and suggested that substrate binding or dimerization may be necessary to force the arrangement of a functional catalytic triad (6, 43). In the Chiba virus protease, Glu 54 is within the hydrogen bonding distance, in two of the four molecules in the crystallographic asymmetric unit, from the His 30, although in the *anti* conformation (29). Other evidence that appears to favor the dyad hypothesis is the mutational analysis of the Chiba virus protease, in which mutation of Glu 54 to Ala does not inactivate the protease activity. However, protease activity in these studies was assessed at a

single time point in *E. coli* lysates (39). More quantitative analysis of the Chiba virus protease may be necessary to establish the role of Glu 54 during proteolysis.

Conformational changes in the catalytic residues in the presence of AEBSF. To obtain more information about the mechanism of proteolysis, we cocrystallized NV^{PRO} in the presence of AEBSF, which irreversibly inhibits serine proteases and is a soluble, stable, nontoxic alternative to PMSF. AEBSF has been shown previously (and in data not shown) to inhibit NV^{PRO} (9) and is suggested to inhibit proteases by the same mechanism as other sulfonyl fluoride inhibitors, such as PMSF, which rapidly react with the active-site serine or cysteine forming a stable acylated enzyme that blocks proteolytic activity. Our structural analysis of the NV^{PRO}, from the crystals grown in the presence of AEBSF, shows that the catalytic residues undergo a significant conformational change. First, the Glu and His residues have shifted away from each other (Fig. 4B), preventing the hydrogen bond formation observed in the native NV^{PRO} structure (Glu 54 Oε2 to His 30 Nδ1 is now 3.4 Å and Glu 54 Oε1 to His 30 N is now 3.7 Å). Also, the water molecule hydrogen bonded to Glu54 Oε2 is no longer present. This movement of the Glu side chain appears to be associated with a conformational change in the side chain of Arg 112, which is implicated in the substrate specificity pocket (see below). This residue, which is directed away from the active site in the native structure, now points toward Glu 54. Second, the side chain of the Cys 139 exhibits two alternate conformations, one of which points away from the His 30, in contrast to the native structure that shows a single conformation pointing toward the His 30. Third, the electron density map shows an unusual flattened, bulged density for His 30 (Fig. 4B) compared to other His residues, which are well defined throughout the rest of the structure and in the native structure.

Although conformational changes in the catalytic residues are observed, the cause of these changes remains unclear due to the lack of electron density corresponding to AEBSF. To examine the possibility of covalent modification by AEBSF, we carried out electrospray mass spectrometry on AEBSF-treated NV^{PRO}; however, this failed to indicate any covalent modification of NV^{PRO}. One possibility is that AEBSF may noncovalently interact with NV^{PRO} either by directly affecting the electrophilic nature of the active site residues or by affecting the conformation of a surface residue such as Arg 112 to disrupt the hydrogen bonding interactions in the active site residues. The absence of AEBSF in the electron density may be due to disorder or low occupancy. Another distinct possibility, is that the addition of AEBSF, an acidic compound, during crystallization may have caused a subtle change in pH resulting in the disruption of hydrogen bond interactions due to protonation of the Glu 54 side chain (pKa ~4.4). This disruption of hydrogen bond interactions may make His 30 more flexible and account for the unusual density; such flexibility could lead to the alternate conformations seen in the side chain of Cys.

Oxyanion hole. Oxyanion holes help substrates bind tightly and are used to stabilize the tetrahedral transition state formation through hydrogen bonding with the negatively charged carbonyl oxygen of the substrate backbone at the cleavage site. The standard oxyanion hole includes the amino acid sequence Gly-Xaa-Ser/Cys-Gly, where serine or cysteine is the active-site nucleophile. The oxyanion hole of NV^{PRO} protease is formed by the sequence Gly 137-Asp 138-Cys 139-Gly 140 (Fig. 5, dark blue). In the serine proteases, the hydroxyl group (OH) of the active site serine is positioned appropriately to make hydrogen bonding interactions with the Nδ2 of the imidazole ring of the active-site histidine. The SH group of the cysteine in the viral cysteine proteases is positioned similarly. However, as seen in our NV^{PRO} structure, the S · · · N distance in the structures of these proteases is generally ca. 3.5 Å (Fig. 4). Perhaps due to the increased electronegativity of the sulfur atom, the cysteine residue is pulled back relative to the corresponding serine residue in serine proteases to achieve the appropriate electrophilic environment conducive for proteolysis.

The sequence that forms the oxyanion hole is highly conserved among the serine and serine-like proteases where aspartic acid occupies the Xaa site in most eukaryotic proteases, bacterial serine proteases, viral 2A cysteine proteases, and the norovirus proteases. However, the viral 3C and 3C-like proteases have a variety of residues present at the Xaa site, including Asp, Gln, Met, Tyr, or Trp (29).

Substrate binding sites. NV^{PRO} cleaves the ORF1 encoded polyprotein at five sites to release the six nonstructural proteins for replication. Although there are five sites, there are only three different types of cleavage sites. There is always a Glu or Gln residue in the N-terminal side of the cleavage site (P1 position) and an Ala or Gly in the C-terminal side (P1′ position) of NV^{PRO} substrates. The Gln-Gly junctions (between p48/p41 and p41/3A) are cleaved from the polyprotein first, followed by the two Glu-Gly junctions (between 3A/Vpg and protease/polymerase) and the Glu-Ala junction (between the VPg/protease) (9). Based on the recently published Chiba virus protease structure modeled with its substrate, and multiple sequence alignments with other calicivirus and picornavirus 3C proteases, Nakamura et al. suggested five residues that may be important for substrate specificity: His 157, Ala 160, Ile 109, Arg 112, and Val 114 (29). These residues are classified, according to the standard convention, as the S1 pocket containing the first two residues, along with several hydrophobic residues (Fig. 5, yellow), and the S2 pocket containing the latter three residues (Fig. 5, cyan). The identified substrate-binding pocket is in general agreement with the crystal structures of the HRV 3C protease with its inhibitors (23). The substrate specificity sites and the active site are located at the cleft between the N-terminal β-sheet and C-terminal β-barrel domains. The putative substrate binding residues of Chiba virus protease are highly conserved among the noroviruses and likely represent the NV^{PRO} binding sites.

S1 pocket and the role of His 157. The position of His 157 (Fig. 5, magenta) in the S1 pocket is conserved structurally, not only between the NV and Chiba virus proteases but also between various picornaviral 3C proteases. In the 1.5-Å structure of the NV^{PRO}, the imidazole of His 157 hydrogen bonds with the side chain of the Thr 134 via a bound water molecule and the side chain of the Tyr 143. All of these residues are conserved throughout the *Caliciviridae* family. It is interesting that similar interactions between conserved His, Thr, and a Tyr are also observed in picornaviral 3C protease structures (6, 7, 24, 27). Mutational analyses of His 157 in both of these virus families have shown that this residue is critically important for substrate binding and cleavage (8, 39). The role of the ob-

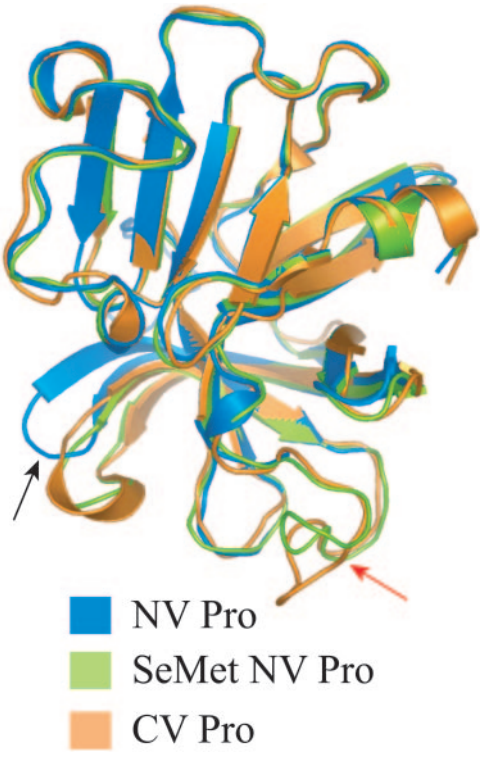

FIG. 6. Comparison of NV^{PRO} and Chiba virus protease. The average RMS deviation between the matching C_α atoms, of each monomer compared to the 1.5-Å NV^{PRO}, is ~0.7 Å. The Norwalk virus protease has a large flexible loop composed of residues 121 to 136 (red arrow) that can adopt several different conformations. This loop is disordered in the 1.5-Å structure of the NV^{PRO} (blue), suggesting that it is very flexible. The loop is visible in the Se-Met NV^{PRO} (green) and Chiba virus protease (orange) structures in two different conformations without disturbing the β-sheets or catalytic triad. Another conformationally flexible region is in the S2 binding pocket (black arrow). This region has moved 4 to 5 Å between the native NV^{PRO} and the Se-Met NV^{PRO} and the Chiba virus protease.

served hydrogen bond interactions may be in order to position His 157 appropriately for substrate binding and cleavage. The other residue in the S1 pocket, Ala 160, is a part of the eII β-strand with its amide hydrogen bonded to the carbonyl of Val 167 in the fII β-strand. These two residues are surrounded by other hydrophobic residues, such as Leu 121, Val 99, Leu 97, Ile 87, and Val 57, to form the hydrophobic S1 pocket. The positions of these residues and the hydrogen bonding pattern involving His 157 are maintained in the AEBSF-NV^{PRO} structure.

S2 pocket and conformational flexibility. The three residues—Ile 109, Arg 112, and Val 114—in the S2 binding pocket are part of the eII and fII β-strands. This region in the 1.5-Å native structure exhibits a significant conformational change of approximately 4 to 5 Å compared to either the Se-Met NV^{PRO} structure or the Chiba virus protease structure (Fig. 6, black arrow). This region is in close proximity to the dimeric interface. Such a conformational change may indeed be due to the fact that in the 1.5-Å structure the dimer is related by a strict crystallographic twofold axis as opposed to the dimers in the other two structures, which are related by an approximate noncrystallographic twofold axis. The Gly residues at positions 102 and 119 are likely involved in conferring the observed conformational flexibility to this region.

In addition to the structural change in the S2 region, comparison of the native NV^{PRO}, AEBSF-NV^{PRO}, Se-Met NV^{PRO}, and Chiba virus protease revealed another significant change in the loop formed by residues 121 to 136 (Fig. 6, red arrow). This loop adopts different conformations in the Se-Met NV^{PRO} and the Chiba virus protease structures and is disordered (aa 122 to 133) in the 1.5- and 2.2-Å NV^{PRO} structures. Even though this loop is in several different conformations, there is little change in the cII or dII β-strands that are connected by this loop. It is interesting that this loop is conserved among noroviruses with the exception of Gly 129 and Ile 136. The different conformations of this loop and the loop in the S2 region, although they may be due to crystal packing or dimeric interactions, underscore the flexible nature inherent in these loops. Such flexibility may be useful in the overall function of NV^{PRO}, which involves cleavage of the polyprotein in five different places.

Conclusions. The proteases of viruses that encode polyproteins are especially important targets for antiviral compounds. We have presented the X-ray structure of the NV protease determined to a 1.5-Å resolution, which provides an excellent structural target for the rational design of protease-specific inhibitors and for mutational studies to further our understanding of the structure-based function of this protein. The overall fold of this protease resembles that of the serine and other viral cysteine proteases; however, the difference is that there is only one β-barrel domain, with the other domain being a slightly twisted β-sheet. Our biochemical and structural studies provide strong evidence for the active role of Glu 54 in proteolysis. From our structural analysis, which is consistent with size exclusion chromatography, NV^{PRO} is a dimer in both solution and the crystal structure; however, the functional significance of dimer formation remains to be established since each monomer contains an independent catalytic triad.

ACKNOWLEDGMENTS

This study was supported by grants from the NIH (PO1 AI057788 to M.K.E. and B.V.V.P.) and the Robert Welch Foundation (to B.V.V.P.). We acknowledge the use of the SBC-CAT 19ID beamline and its staff for help during data collection at the Advanced Photon Source supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38.

We thank Richard Cook for help with fluorogenic peptide synthesis and mass spectrometry and Theodore Wensel for helpful discussions on FRET analysis.

REFERENCES

1. **Allaire, M., M. M. Chernaia, B. A. Malcolm, and M. N. James.** 1994. Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature **369:**72–76.
2. **Anand, K., J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld.** 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science **300:**1763–1767.
3. **Babe, L. M., and C. S. Craik.** 1997. Viral proteases: evolution of diverse structural motifs to optimize function. Cell **91:**427–430.
4. **Bazan, J. F., and R. J. Fletterick.** 1988. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. Proc. Natl. Acad. Sci. USA **85:**7872–7876.
5. **Belliot, G., S. V. Sosnovtsev, T. Mitra, C. Hammer, M. Garfield, and K. Y. Green.** 2003. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. J. Virol. **77:**10957–10974.
6. **Bergmann, E. M., S. C. Mosimann, M. M. Chernaia, B. A. Malcolm, and M. N. James.** 1997. The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. J. Virol. **71:**2436–2448.

7. **Birtley, J. R., S. R. Knox, A. M. Jaulent, P. Brick, R. J. Leatherbarrow, and S. Curry.** 2005. Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. J. Biol. Chem. **280:**11520–11527.
8. **Blair, W. S., J. H. Nguyen, T. B. Parsley, and B. L. Semler.** 1996. Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. Virology **218:**1–13.
9. **Blakeney, S. J., A. Cahill, and P. A. Reilly.** 2003. Processing of Norwalk virus nonstructural proteins by a 3C-like cysteine proteinase. Virology **308:**216–224.
10. **Boniotti, B., C. Wirblich, M. Sibilia, G. Meyers, H. J. Thiel, and C. Rossi.** 1994. Identification and characterization of a 3C-like protease from rabbit hemorrhagic disease virus, a calicivirus. J. Virol. **68:**6487–6495.
11. **Bricogne, G., C. Vonrhein, C. Flensburg, M. Schiltz, and W. Paciorek.** 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. **59:**2023–2030.
12. **Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallographica Sect. D **54:**905–921.
13. **Daughenbaugh, K. F., C. S. Fraser, J. W. Hershey, and M. E. Hardy.** 2003. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. EMBO J. **22:**2852–2859.
14. **DeLano, W. L.** 2002. The PyMOL molecular graphics system. DeLano Scientific, San Carlos, Calif.
15. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60:**2126–2132.
16. **Ettayebi, K., and M. E. Hardy.** 2003. Norwalk virus nonstructural protein p48 forms a complex with the SNARE regulator VAP-A and prevents cell surface expression of vesicular stomatitis virus G protein. J. Virol. **77:**11790–11797.
17. **Glass, P. J., L. J. White, J. M. Ball, I. Leparc-Goffart, M. E. Hardy, and M. K. Estes.** 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. J. Virol. **74:**6581–6591.
18. **Hardy, M. E., T. J. Crone, J. E. Brower, and K. Ettayebi.** 2002. Substrate specificity of the Norwalk virus 3C-like proteinase. Virus Res. **89:**29–39.
19. **Jiang, X., M. Wang, D. Y. Graham, and M. K. Estes.** 1992. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. J. Virol. **66:**6527–6532.
20. **Jones, S., and J. M. Thornton.** 1996. Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA **93:**13–20.
21. **Jones, T. A., J. Y. Zou, S. W. Cowan, and Kjeldgaard.** 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47**(Pt. 2)**:**110–119.
22. **Laskowski, R. A., D. S. Moss, and J. M. Thornton.** 1993. Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. **231:**1049–1067.
23. **Matthews, D. A., P. S. Dragovich, S. E. Webber, S. A. Fuhrman, A. K. Patick, L. S. Zalman, T. F. Hendrickson, R. A. Love, T. J. Prins, J. T. Marakovits, R. Zhou, J. Tikhe, C. E. Ford, J. W. Meador, R. A. Ferre, E. L. Brown, S. L. Binford, M. A. Brothers, D. M. DeLisle, and S. T. Worland.** 1999. Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc. Natl. Acad. Sci. USA **96:**11000–11007.
24. **Matthews, D. A., W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E. Villafranca, C. A. Janson, H. E. McElroy, C. L. Gribskov, et al.** 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. Cell **77:**761–771.
25. **McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read.** 2005. Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. **61:**458–464.
26. **Miller, R., and S. M. Gallo.** 1994. *SnB*: crystal structure determination *via* shake-and-bake. J. Appl. Crystallogr. **27:**613–621.
27. **Mosimann, S. C., M. M. Cherney, S. Sia, S. Plotch, and M. N. James.** 1997. Refined X-ray crystallographic structure of the poliovirus 3C gene product. J. Mol. Biol. **273:**1032–1047.
28. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. **53:**240–255.
29. **Nakamura, K., Y. Someya, T. Kumasaka, G. Ueno, M. Yamamoto, T. Sato, N. Takeda, T. Miyamura, and N. Tanaka.** 2005. A norovirus protease structure provides insights into active and substrate binding site integrity. J. Virol. **79:**13685–13693.
30. **Ng, K. K., N. Pendas-Franco, J. Rojo, J. A. Boga, A. Machin, J. M. Alonso, and F. Parra.** 2004. Crystal structure of Norwalk virus polymerase reveals the carboxyl terminus in the active site cleft. J. Biol. Chem. **279:**16638–16645.
31. **Otto, H. H., and T. Schirmeister.** 1997. Cysteine proteases and their inhibitors. Chem. Rev. **97:**133–172.
32. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:**307–326.
33. **Perrakis, A., M. Harkiolaki, K. S. Wilson, and V. S. Lamzin.** 2001. ARP/wARP and molecular replacement. Acta Crystallogr. D Biol. Crystallogr. **57:**1445–1450.
34. **Peters, H., Y. Y. Kusov, S. Meyer, A. J. Benie, E. Bauml, M. Wolff, C. Rademacher, T. Peters, and V. Gauss-Muller.** 2005. Hepatitis A virus proteinase 3C binding to viral RNA: correlation with substrate binding and enzyme dimerization. Biochem. J. **385:**363–370.
35. **Pfister, T., and E. Wimmer.** 2001. Polypeptide p41 of a Norwalk-like virus is a nucleic acid-independent nucleoside triphosphatase. J. Virol. **75:**1611–1619.
36. **Pflugrath, J. W.** 1999. The finer things in X-ray diffraction data collection. Acta Crystallogr. D Biol. Crystallogr. **55**(Pt. 10)**:**1718–1725.
37. **Seipelt, J., A. Guarne, E. Bergmann, M. James, W. Sommergruber, I. Fita, and T. Skern.** 1999. The structures of picornaviral proteinases. Virus Res. **62:**159–168.
38. **Someya, Y., N. Takeda, and T. Miyamura.** 2005. Characterization of the norovirus 3C-like protease. Virus Res. **110:**91–97.
39. **Someya, Y., N. Takeda, and T. Miyamura.** 2002. Identification of active-site amino acid residues in the Chiba virus 3C-like protease. J. Virol. **76:**5949–5958.
40. **Sosnovtsev, S. V., M. Garfield, and K. Y. Green.** 2002. Processing map and essential cleavage sites of the nonstructural polyprotein encoded by ORF1 of the feline calicivirus genome. J. Virol. **76:**7060–7072.
41. **Wirblich, C., M. Sibilia, M. B. Boniotti, C. Rossi, H. J. Thiel, and G. Meyers.** 1995. 3C-like protease of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity. J. Virol. **69:**7159–7168.
42. **Yennawar, N. H., H. P. Yennawar, and G. K. Farber.** 1994. X-ray crystal structure of gamma-chymotrypsin in hexane. Biochemistry **33:**7326–7336.
43. **Yin, J., E. M. Bergmann, M. M. Cherney, M. S. Lall, R. P. Jain, J. C. Vederas, and M. N. James.** 2005. Dual modes of modification of hepatitis A virus 3C protease by a serine-derived beta-lactone: selective crystallization and formation of a functional catalytic triad in the active site. J. Mol. Biol. **354:**854–871.